Legislative analysis

Mokken, Robert J.; Stokman, Frans N.

COALITION FORMATION
Henk A.M. Wilke (ed.)

Chapter 4

LEGISLATIVE ANALYSIS: METHODOLOGY FOR THE ANALYSIS OF GROUPS AND COALITIONS

Robert J. Mokken and Frans N. Stokman

In this chapter a theory and methods are given that enable us to analyze relations within and between groups and coalitions in decision-making bodies on the basis of actual behavior of the individual group members. The theory is suitable for the analysis of roll calls, interview questions with three response categories and preference rank orders. The methods and coefficients based on it provide large possibilites to company policy positions of decision-makers with respect to each other, to determine policy location and cohesion of groups of decision-makers, to indicate the location of group members relative to the group, and to search for blocs or cliques of decision-makers in terms of a certain minimal level of cohesion. In a wider context the theory and methods might well be used to investigate the degree of consensus among political, social, and economic elites regarding basic values and norms; or they can be used in experimental designs to analyze the consequences of different independent variables on the cohesion of small groups.

Contents

Introduction

Mokken and Stokman (1970) demonstrated that the types of three-valued data that confront us in roll call situations (yes-abstention-no) have certain metric properties. From some mild assumptions a simple measure of the distance between pairs of decision-makers was derived on the basis of a pairwise comparison of their voting behaviour on a given set of issues (roll calls). The analytic possibilities of this distance measure have been demonstrated in the study of Stokman (1977) on Third World Group information in the United Nations.[1)] In section 4.1 we shall give the distance metric and the assumptions on which it is based. That distance metric can be normalized in two ways: on the basis of its maximum value and on the basis of the expected distance between pairs of decision-makers in case of random voting. We shall consider these two possibilities in section 4.1 and choose the second.

Using a related system of axioms Kemeny and Snell (1962) showed that a similar distance can be defined between decision-makers or more general, persons, in terms of their expressed preference order concerning a common set of items. The whole conceptual and procedural framework of our distance analysis can therefore be extended to such preferential data (Hazewindus and Mokken, 1972). In section 4.2 this extention to preference rank orders will be given.

In section 4.3 a number of measures will be put forward for group analysis, the analysis of a priori defined groups of decision-makers (e.g. factions in parliament, caucusing groups in the U.N. General Assembly). For the evaluation of possible coalitions two characteristics of groups of decision-makers seem to be of outmost importance: the policy locations of groups with respect to each other and the cohesiveness of the groups. For both characteristics measures will be proposed. These measures are related to characteristics of groups as a whole. Other measures will be proposed to evaluate the policy location of individual decision-makers with respect to its own and other groups. The analytic possibilities of these measures will be extensively illustrated in two completely different settings.

For the analysis of cohesiveness of groups and coalitions two coefficents of cohesion will be proposed, one based on the average distance in the group and one based on the largest distance in the group. These two coefficients can also be used for procedures of bloc- and clique analysis that enable us to detect informal groups or coalitions. In these two kinds of analysis no groups are a priori defined: blocs or cliques of decision-makers are defined in terms of a certain minimal level of cohesion, observed in their votes or preferential orders. We shall give in section 4.4 a procedure of bloc analysis, based on the average distance in the bloc, and in section 4.5 a procedure of clique analysis, based on the maximum distance in the cluster. In section 4.6 we shall consider two possible solutions to the problem of nonvoting due to absence: the treatment of absence as abstention and the treatment of absence as missing data. Arguments for and consequences of both will be given. Finally, in section 4.7 our theory and methods will be compared with other methods of roll call analysis and analysis of preferential rank orders.

4.1 The distance metric in roll call situations.

To determine policy locations and cohesion of groups of decision-makers a pairwise comparison will be made of their voting behaviour in roll calls (or for that matter of their responses on 'voting-like' questions in questionnaires). Usually a decison-maker has three voting alternatives in a roll call: he can vote in favour, against, or he can abstain.[2)] For the moment we identify absence (non-voting) and abstention as one outcome of voting. We do not know the intensity of the vote: one decision-maker may strongly be in favour, whereas another decision-maker may only be reaching a minimal compromise. Nor do we know the reasons of the decision-maker for his voting bahaviour. A decision-maker may vote in a certain way, because he perceives that voting alternative as the best for his own interests, or because he agreed to choose that alternative in a logrolling situation with another decision-maker.

Let us formalize the voting situation, for convenience, in the following way. We consider a body of m decision-makers, voting on n roll calls. Roll calls are designated by the index i:1, 1, ..., n. Decision-makers are denoted by A, B, C, ... and their votes on roll call i by the corresponding lowercase symbols a_i, b_i, c_i, ... The three voting alternatives are described as follows:

decision-maker A votes in favour or yes : $a_i = y$
decision-maker A votes against or no : $a_i = n$
decision-maker A abstains : $a_i = ?$

In table 4.1 we give hypothetical outcomes for five decision-makers voting on five roll calls.

Each decision-maker is characterized by the set of votes cast by him on the five roll calls. For decision-maker A this set is given as:

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} \text{ or } \begin{pmatrix} y \\ y \\ n \\ n \\ y \end{pmatrix}$$

In general we call the set

$$\vec{a} = \begin{pmatrix} a_1 \\ . \\ . \\ . \\ . \\ a_n \end{pmatrix}$$

the <u>vote array</u> of decision-maker A on n roll calls. The <u>vote vector</u> $\vec{a}$ of decision-maker A is a numerical representation of his vote array. It is obtained by a <u>function</u> s, which prescribes for each element a_i of the vote array a corresponding numerical value $s(a_i)$ in the vote vector. An example of a numerical representation of the vote array of decision-maker B in table 4.1 is the following vote vector:

$$\begin{pmatrix} y \\ n \\ n \\ ? \\ y \end{pmatrix} \xrightarrow{\quad s \quad} \begin{pmatrix} 1 \\ -1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

Let us consider in table 4.1 the vote arrays $\vec{a}$, $\vec{b}$, and $\vec{e}$ of decision-makers A, B, and E. It can be seen that decision-makers A and E opposed each other on all roll calls except roll call 3 where they agreed: $a_3 = e_3 = n$. Considering the vote array $\vec{b}$ of decision-maker B, we note that B voted with A or E on all roll calls except roll call 4. On roll call 4 B voted abstention ($b_4 = ?$). On that roll call his position can be considered as '<u>between</u>' that of decision-maker A ($a_4 = n$) and E ($e_4 = y$): B neither supported nor opposed the proposal. For the whole vote array $\vec{b}$ of decision-maker B we may conclude that B took either A's or E's part or voted in

Decision-makers:		A ($\vec{a}$)	B ($\vec{b}$)	C ($\vec{c}$)	D ($\vec{d}$)	E ($\vec{e}$)
roll calls:	1	y	y	y	y	n
	2	y	n	?	n	n
	3	n	n	y	n	n
	4	n	?	y	?	y
	5	y	y	?	?	n

Table 4.1

Vote arrays of five decision-makers voting on five roll calls.

between. Consequently, we may reason that B's voting position on the particular set of roll calls is between that of A and E. The reader should note that if B had voted yes ($b_3 = y$) or abstention ($b_3 = ?$) on roll call 3, there would be no reason to consider B's vote position to be between that of A and E on this set of roll calls. A close scrunity of table 4.1 reveals that decision-maker C is not between any other pair of decision-makers (see roll call 3, where $c_3 = y$ and all other decision-makers voted no). Therefore, the concept of 'betweenness' is introduced in two ways: first, the vote alternative abstention ($a_i = ?$) indicates a policy position between the policy positions expressed by in favor ($a_i = y$) and against ($a_i = n$); second, betweenness is a characteristic of a vote array: over a set of roll calls the policy position of a decision-maker B is between those of A and C, if for all roll calls in which A and C chose the same vote alternative B also chose that alternative ($a_i = b_i = c_i$) and for all roll calls in which A and C had different votes ($a_i \neq c_i$) B either voted as A ($b_i = a_i$) or as C ($b_i = c_i$) or abstained ($b_i = ?$).

Another important concept in the axiomatic development of the distance metric is the converse of a vote. We may ask how a decision-maker would have voted, if the proposal was negatively formulated. In our opinion an adequate negative formulation of the proposal would have led yes voters to vote no, no voters to vote yes, and abstain voters still to abstain. If we denote the converse of a vote a_i on a roll call i by $-a_i$, we have therefore:

$$
\begin{array}{ll}
a_i = y & -a_i = n \\
a_i = n & -a_i = y \\
a_i = ? & -a_i = ?
\end{array}
$$

This definition of the converse of a vote can now be used to define the converse of a vote array $\vec{a}$. A decision-maker B is said to be the converse of A if his vote on all roll calls is the converse of the vote A: if A voted in favor, B voted against, and vice versa, or they both abstained. Thus a vote array $\vec{b}$ is the converse of a vote array $\vec{a}$, if for all roll calls $i : b_i = -a_i$. We then denote $\vec{b}$ by $-\vec{a}$.

4.1.1 The distance function.

Mokken and Stokman (1970) listed in the form of axioms a number of desirable properties that we desired for our method of measurement. They showed that one and only one distance function exists that satisfies these axioms.

The reader should be aware that once he acknowledges the plausibility of the desirable properties, his acceptance of them as axioms implies that there is only this one distance measure that he can use without contradicting his own assumptions. In the following we will elucidate these basic axioms.

A measure of the distance between two decision-makers A and B is a function on pairs of vote vectors $\vec{a}$ and $\vec{b}$ to be denoted by $d(\vec{a},\vec{b})$.

First we want our distance measure to have the three usual properties for all distance functions:

1. Nonnegativity. $d(\vec{a},\vec{b}) \geq 0$, with equality if and only if $\vec{a} = \vec{b}$. The distance between two decision-makers A and B should be a nonnegative number. The distance is zero, if and only if A and B vote alike on all roll calls.
2. Symmetry. $d(\vec{a},\vec{b}) = d(\vec{b},\vec{a})$. The distance from $\vec{a}$ to $\vec{b}$ should be the same as that from $\vec{b}$ to $\vec{a}$; i.e. the distance between pairs of votes should not depend on the order of comparison.
3. Trangle inequality. For any three vote vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$

   $d(\vec{a},\vec{b}) + d(\vec{b},\vec{c}) \geq d(\vec{a},\vec{c})$

   with equality if and only if the vote vector $\vec{b}$ is 'between' the vote vectors $\vec{a}$ and $\vec{c}$. The reader should realize that this third axiom implies that we need not investigate the whole vote vector on n roll calls for each of the three decision-makers A, B and C in order to decide whether B is between A and C (see table 4.1). It is sufficient to compare just three numbers: $d(\vec{a},\vec{b})$, $d(\vec{a},\vec{c})$ and $d(\vec{b},\vec{c})$ to establish this. This axiom has therefore some analytic possibilities.

   The fourth axiom or desirable property of our distance function $d(\vec{a},\vec{b})$ concerns the order of roll calls. We do not want the distance between two decision-makers A and B to depend on the accidental order in which the roll calls were taken. Permutation of roll calls should not change the distances between decision-makers. The fourth axiom can therefore be denoted as:
4. Invariance for identical permutations. The distance $d(\vec{a},\vec{b})$ is invariant for identical permutation of the roll calls in the two vote vectors $\vec{a}$ and $\vec{b}$.

   Let us now consider the situation that two decision-makers voted alike on all but $k (k \leq n)$ roll calls. Decision-makers A and B in table 4.1, for instance, differ for the five roll calls only on roll calls 2 and 4. In such a case we may then find it natural to require that the distance computed for the two decision-makers A and B over all five roll calls be the same as the distance computed between the two decision-makers for only those two roll calls on which they differ. In other words, the distance between two decision-makers should not change when we add to their vote vector only those roll calls on which they voted alike. Consequently, distances between decision-makers are created only by differing votes. This desirable property is formulated in a fifth axiom:

5. The distance between two vote vectors $d(\vec{a},\vec{b})$ depends only on those roll calls in which A and B voted differently ($a_i \neq b_i$).

In the sixth axiom a unit is specified, in which the distances are to be measured. Given the discrete nature of the data, we postulated:

6. The minimum positive distance is 1.

Finally, we need a very important seventh axiom, stating that the distance between two decision-makers A and B will not change, if the proposals are put to the vote in their negative formulation. So, if we substitute for the vote vector $\vec{a}$ and $\vec{b}$ their converse vectors $-\vec{a}$ and $-\vec{b}$, the value of the distance function does not change:

7. Invariance for taking converses. For any two vote vectors $\vec{a}$ and $\vec{b}$ $d(\vec{a},\vec{b}) = d(-\vec{a},-\vec{b})$.

Let us now consider two decision-makers A and B with their vote arrays $\vec{a}$ and $\vec{b}$. Suppose that they voted alike on all roll calls except the last one, where A abstained ($a_n = ?$) and B voted yes ($b_n = y$). We can readily see that these two decision-makers differ minimally as no vote array $\vec{c}$ can be found in which $\vec{c}$ is between $\vec{a}$ and $\vec{b}$ and $d(\vec{a},\vec{c}) \geq 0$, $d(\vec{c},\vec{d}) \geq 0$. That is, no array $\vec{c}$ can be found that is different from $\vec{a}$ and $\vec{b}$ and yet between them. Therefore, the distance between $\vec{a}$ and $\vec{b}$ must be minimal and by axiom 6 equal to 1. Applying axiom 5, we find: $d(\vec{a},\vec{b}) = d(?,y) = 1$. An application of axiom 7 shows that also: $d(?,y) = d(-?,-y) = d(?,n) = 1$. We can see here that axiom 7 amounts to the assumption that the distance between abstention and yes is the same as the distance between abstention and no. It can now be proven that for one roll call ($n = 1$) all distances are determined. All interpair distances are plotted in table 4.2.

		b: y	$?$	n
	y	0	1	2
a	$?$	1	0	1
	n	2	1	0

Table 4.2

Distance values d(a,b) on single roll calls.

Having established these distances between pairs of decision-makers on a single roll call, we are now able to assign numerical values to the vote alternatives in such a way that these distances are preserved. An easy way to achieve this is to map the vote into the following vote scores:

$$\begin{pmatrix} y \\ ? \\ n \end{pmatrix} \xrightarrow{\ s\ } \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

Table 4.3 contains the vote vectors - the numerical representations - of the vote arrays of table 4.1.

Decision-makers	A $(\vec{a})$	B $(\vec{b})$	C $(\vec{c})$	D $(\vec{d})$	E $(\vec{e})$
roll calls: 1	1	1	1	1	-1
2	1	-1	0	-1	-1
3	-1	-1	1	-1	-1
4	-1	0	1	0	1
5	1	1	0	0	-1

Table 4.3

Vote vectors of table 4.1.

From now on we assume that the voted are scored, in this numerical way. Mokken and Stokman (1970, 3-12, theorem 3.1) proved that for any roll calls all distances are uniquely determined. The unique distance function between two vote vectors $\vec{a}$ and $\vec{b}$, denoted by $d(\vec{a},\vec{b})$ is (Mokken and Stokman, 1970, 3-12, theorem 3.2):

$$d(\vec{a},\vec{b}) = \sum_{i=1}^{n} |a_i - b_i| = \sum_{i=1}^{n} d(a_i,b_i) \qquad (4.1)$$

This city bloc or Manhattan metric (Attneave, 1950; Torgerson, 1958) is used to derive a number of concepts and methods which offer a wide spectrum of possibilities for the type of roll call analysis in which we are interested.[3] The distances of all pairs of decision-makers can be computed by (4.1) and represented in a distance matrix

$$D = \| d(\vec{a},\vec{b}) \| \text{ of order } m \times m \qquad (4.2)$$

where the vote vectors $\vec{a}$ and $\vec{b}$ represent the voted vectors of any pair of decision-makers A and B in any set of m decision-makers we consider.

D is a symmetric matrix with zero elements on the principal diagonal. For our example of tables 4.1 and 4.3 this matrix is given in table 4.4.

	A	B	C	D	E
A	0	3	6	4	8
B	3	0	5	1	5
C	6	5	0	4	6
D	4	1	4	0	4
E	8	5	6	4	0

Table 4.4.
Distance matrix for tables 4.1 and 4.3.

4.1.2 Normalized distance.

Distances as defined above depend on the numer n of roll calls on which votes are compared. The distance between two decision-makers therefore is also a function of the number of votes in that sense that for a large number of votes any two decision-makers can diverge more widely in terms of large distances than for a small number of votes. We can easily see that the maximum possible distance for a set of n issues is:

$$d_{max} = 2n.$$

Sometimes it may be necessary to relate the observed distance between a pair of decision-makers to that maximum distance. For instance, if we want to compare the distances of such a given pair of decision-makers over different periods of analysis, we should take into account that the sets may contain different numbers of roll calls. Consequently, the maximum possible differences may vary across such sets. In these cases we may want to normalize the distances for each set of roll calls in such a way that d_{max} is constant across sets. We achieve this with the following definition: a normalized distance between two vote vectors $\vec{a}$ and $\vec{b}$ on a set of n roll calls is given by

$$d^*(\vec{a},\vec{b}) = \frac{d(\vec{a},\vec{b})}{d_{max}} = \frac{d(\vec{a},\vec{b})}{2n} \qquad (4.3)$$

Obviously, we have

$$0 < d^*(\vec{a},\vec{b}) < 1$$

for any set of roll calls.[4)]

Corresponding to d^* and (4.2) we define the normalized distance matrix.

$$D^* = \| d^*(a,b) \| \qquad (4.4)$$

4.1.3 Reduced distance.

In the last section we related the distance between a pair of decision-makers on n roll calls to the maximum distance over that set of n roll calls ($2n$). We must still judge whether a certain value of the normalized distance, e.g. .5(d^* = .5), should be considered as a small or a large distance between two decision-makers. The normalized distance does not provide a yardstick against which to measure the actual distance of a pair of decision-makers. A criterion, however, might be found in the expected distance between them in the hypothetical case of random voting. What is the meaning of such a criterion? Of course, we expect that decision-makers do not vote randomly in real-life situations. If we find an actual distance that is approximately equal to the expected distance in the case of random voting, we do not conclude that these two decision-makers actually voted randomly. However we assume that two decision-makers with rather similar policy positions will have a much smaller distance than that expected in the case of random voting. We also assume that two decision-makers with strongly different policy positions will have a much larger distance than that expected in the case of random voting. The criterion of the expected distance therefore provides a yardstick for judging whether the actual distance expresses similar or dissimilar policy positions. Moreover, the hypothesis of random voting enables us to develop a statistical theory of roll call analysis, which considerably adds to its value as a criterion. It enables us, as we shall see in the next sections, to give statistical criteria by which to judge whether coefficients based on our distance function are relatively high or low.

For the determination of the expected distance and its sampling distribution we assume statistical independence between the roll calls. This condition is unrealistic in the parliamentary setting of compromises and log-rolling, and does not mean that there is no systematic variation over roll calls in reality. It is only introduced to construct a yardstick against which the actual distances can be judged. All systematic deviation from the null case of random and independent voting in real-life situations is then considered to be caused by the respective policy positions of the decision-makers.

If the expected distance of a pair of decision-makers A and B in case of random and mutually independent voting is denoted by $\mathcal{E}\{\underline{d}(\vec{a},\vec{b})\}$, the following definition can be given:[5)]

The reduced distance between two vote vectors $\vec{a}$ and $\vec{b}$ on a set of n issues is given by

$$d^r(\vec{a},\vec{b}) = \frac{d(\vec{a},\vec{b})}{\mathcal{E}\{\underline{d}(\vec{a},\vec{b})\}} \qquad (4.5)$$

It should be noted that the reduced distances still have all the properties of a distance, because they are obtained by dividing the absolute distances by a positive constant. Thus, the reduced distance is obtained from the absolute distance only by a scale transformation that does not change the character of the distance function. The corresponding reduced distance matrix is defined as:

$$D^r = \| d^r(\vec{a},\vec{b}) \| \qquad (4.6)$$

The value of the expected distance $\mathcal{E}\{\underline{d}(\vec{a},\vec{b})\}$ depends on the number of issues and the probabilities of voting in favor, voting against, and abstention. We shall consider the expected distance between two decision-makers A and B first on one roll call, and then on n roll calls.

For any decision-maker A the outcome of its vote a on a roll call is determined by the following probabilities:

$$\begin{aligned} p(a_1) &= \pi_1 = P\{\underline{a} = 1\} \\ p(a_0) &= \pi_0 = P\{\underline{a} = 0\} \\ p(a_{-1}) &= \pi_{-1} = P\{\underline{a} = -1\} \\ \pi_1 + \pi_0 &+ \pi_{-1} = 1 \end{aligned} \qquad (4.7)$$

Thus, we suppose that on each roll call A votes in favor with the probability π_1, against with the probability π_{-1}, and abstention with the probability π_0. For example, we suppose that A chooses all alternatives with the same probability: $\pi_1 = \pi_0 = \pi_{-1} = 1/3$. Other probabilities are possible, however, and, as we shall see, sometimes even more adequate. We are now able to give the probability space of the joint outcomes of the votes of A and B on a roll call. This is given in table 4.5. Each cell contains the probability of that joint outcome with the corresponding distance between A and B.

		b		
		+1	0	-1
	+1	$\pi_1\pi_1$ d=0	$\pi_1\pi_0$ d=1	$\pi_1\pi_{-1}$ d=2
a	0	$\pi_0\pi_1$ d=1	$\pi_0\pi_0$ d=0	$\pi_0\pi_{-1}$ d=1
	-1	$\pi_{-1}\pi_1$ d=2	$\pi_{-1}\pi_0$ d=1	$\pi_{-1}\pi_{-1}$ d=0

Table 4.5.
Probability space of the joint outcomes
of the vote of A and B on a roll call.

From table 4.5 we can compute the probabilities of the different values of the distance between A and B:

$$\begin{aligned} P\{\underline{d}(a,b) = 0\} &= \pi_1\pi_1 + \pi_0\pi_0 + \pi_{-1}\pi_{-1} \\ P\{\underline{d}(a,b) = 1\} &= 2\pi_0\pi_1 + 2\pi_0\pi_{-1} \\ P\{\underline{d}(a,b) = 2\} &= 2\pi_{-1}\pi_1 \end{aligned} \qquad (4.8)$$

Applying the definition of the expected value, we derive the expected distance as following:

$$\mathcal{E}(\underline{d}) = \mathcal{E}\{\underline{d}(a,b)\} = \Sigma_{d=0}^{2} d\mathrm{P}(\underline{d})$$
$$= 2\pi_0\pi_1 + 2\pi_0\pi_{-1} + 4\pi_1\pi_{-1}$$
$$= \pi_1(\pi_0+2\pi_{-1}) + \pi_0(\pi_1+\pi_{-1}) + \pi_{-1}(2\pi_1 + \pi_0) \quad (4.9)$$

Until now, we have been considering only one roll call. In the definition of the distance over n roll calls in (4.1) we saw that:

$$d(\vec{a},\vec{b}) = \Sigma_{i=1}^{n} d(a_i,b_i)$$

where the $d(a_i,b_i)$ are identically distributed with:

$$\mathcal{E}(\underline{d}) = \mathcal{E}\{\underline{d}(a,b)\} = \mathcal{E}\{\underline{d}(a_i,b_i)\}$$

The expected distance over n roll calls is therefore n times the expected distance on one roll call

$$\mathcal{E}\{\underline{d}(\vec{a},\vec{b})\} = n\mathcal{E}(\underline{d}) \quad (4.10)$$

From the definition of $d^r(\vec{a},\vec{b})$ in (4.5) and the expected distance over n roll calls in (4.10) we can rewrite definition (4.5) as follows:

$$d^r(\vec{a},\vec{b}) = \frac{d(\vec{a},\vec{b})}{n\mathcal{E}(\underline{d})} = \frac{\frac{1}{n}\Sigma_{i=1}^{n} d(a_i,b_i)}{\mathcal{E}(\underline{d})} \quad (4.11)$$

What can we say about this reduced distance, if we assume random and independent voting? It will be evident that its expected value in the case of random voting is 1. Because of the particular form of the reduced distance, as given in (4.11), we are able to derive its asymptotic distribution. In (4.11) the reduced distance was written as a mean distance over n roll calls, divided by a constant $\mathcal{E}(\underline{d})$. This enables us to apply a limit theorem, the Lindeberg-Levy Central Limit Theorem (Rao, 1973, 127), to prove that <u>the reduced distance is approximately normally distributed for large n roll calls</u>.

If the expected distance, and therefore the value of the reduced distance, depends on the probabilities with which the different voting alternatives are being chosen in the case of random voting, which values of π_1, π_0 and π_{-1} should be taken? One solution could be to assume equal probabilities of 1/3 for the three voting alternatives:

$$\pi_1 = \pi_0 = \pi_{-1} = 1/3.$$

In that case the expected distance on one roll call between two decision-makers is:

$$\mathcal{E}(\underline{d}) = \frac{8}{9}$$

We see that a normalized distance of .5 ($d^* = .5$, corresponding with an absolute distance $d = n$) is a larger distance than that expected under random voting conditions with equal probabilities for the voting alternatives. Thus, we are now able to answer the question, whether such a normalized distance should be considered large or not.

The assumption of equal probabilities may not be realistic, however. Generally, in parliaments and in other decision-making bodies most resolutions are adopted. The proportion of votes in favor of a proposal tends to be larger than that against. For this reason it may be more realistic to choose other values for π_1, π_0 and π_{-1} in our model of random voting. For instance, Lijphart (1963, 906-7), following the same argument, suggested $\pi_1 = .5$, $\pi_0 = .2$ and $\pi_{-1} = .3$ for the UN General Assembly. He explicitly stated that in this case the expected agreement would be higher - in our terminology the expected distance smaller - than with equal probabilities for the voting alternatives. Applying definition (4.9), however, we obtain the opposite result:

$$\mathcal{E}(\underline{d} \mid \pi_1 = .5,\ \pi_0 = .2) = .92 > \frac{8}{9} = .89 = \mathcal{E}(\underline{d} \mid \pi_1 = 1/3,\ \pi_0 = 1/3)$$

If we do not want to choose the equal probabilities of 1/3, which probabilities should then be chosen? One way is to estimate the unknown probabilities π_1, π_0 and π_{-1} on the basis of the observed frequencies of voting in favor, against and abstention. In that case π_1, π_0 and π_{-1} are estimated:

$$\hat{\pi}_i = \frac{1}{n} \Sigma_{i=1}^{n} \frac{m_{ij}}{m} \qquad (4.12)$$

In which m_{ij} is the number of decision-makers voting for alternative i ($i = 1, 0, -1$) on roll call j ($j = 1, \ldots, n$), and m is the total number of decision-makers in the decision-making body. $\hat{\pi}_i$ therefore is given as the mean proportion of decision-makers voting for that alternative, taken over all n issues.

Whereas Lijphart assumed that the expected distance in the case of $\pi_1 = .5$, $\pi_0 = .2$ and $\pi_{-1} = .3$ was smaller than in the case of equal probabilities, we have shown the contrary. At which values of π_1, π_0 and π_{-1} is the expected distance maximal? It can be proven that the expected distance $\mathcal{E}(\underline{d})$ between two decision-makers is maximal if:

$$\pi_1 = \pi_{-1} = \tfrac{1}{2} \qquad \pi_0 = 0$$

This indicates the importance of the alternative abstention: the availability of the third alternative abstention is distance-reducing in a decision-making body. The distance-reducing effect of the alternative abstention can be stated even more generally: it can be proven that given a certain proportion between votes in favor and votes against (π_1/π_{-1}), the expected distance increases with decreasing probability π_0. These results can be used to choose a yardstick for the measurement of the cohesion of a whole decision-making body, e.g. the whole UN General Assembly. In section 4.3 we shall propose a coefficient of cohesion, based on the average reduced distance in the group. To judge the cohesion of the whole UN General Assembly we should then use the probabilities $\pi_1 = \pi_{-1} = \frac{1}{2}$, giving the largest expected distance, instead of the equal probabilities $\pi_1 = \pi_0 = \pi_{-1} = 1/3$, as advocated by Willetts (1972, 577), or estimated probabilities.

The reduced distance, which we treated in this section, and the corresponding reduced distance matrix are the basis of our coefficients and procedures, which we shall give in the sections 4.3 to 4.7.[7)]

4.2 The distance metric for preference rank orders.

In this section we will consider preference rank orders between decision-makers to determine policy locations and cohesion of groups of decision-makers. We assume that decision-makers have expressed their preference order for a common set of n items. A, B and C devote decision-makers; $\vec{a}$, $\vec{b}$ and $\vec{c}$ their preference arrays or vectors (cf. section 4.1).

4.2.1 The distance funtion

Kemeny and Snell (1962) proved that, on the basis of six axioms, an uniquely defined distance metric could be obtained between persons in terms of their preference orders for a common set of items. The six axioms of Kemeny and Snell are closely related to the first six axioms given in section 4.1.1.:

1. Nonnegativity. $d(\vec{a},\vec{b}) \geq 0$ with equality if and only if $\vec{a} = \vec{b}$.
2. Symmetry. $d(\vec{a},\vec{b}) = d(\vec{b},\vec{a})$.
3. Triangle inequality. $d(\vec{a},\vec{b}) + d(\vec{b},\vec{c}) \geq d(\vec{a},\vec{c})$ with equality if and only if $\vec{b}$ is 'between' $\vec{a}$ and $\vec{c}$. Preference order $\vec{b}$ is 'between' $\vec{a}$ and $\vec{c}$ if and only if for every pair of items i, j in the set of n items the

ordering of $\vec{b}$ is similar to the ordering of $\vec{a}$ or that of $\vec{c}$ for that pair.

4. Invariance under idential permutations. The distance $d(\vec{a},\vec{b})$ is invariant for identical permutations of the ordered elements of the set of items in the preference vectors $\vec{a}$ and $\vec{b}$.
5. The distance between two preference vectors $d(\vec{a},\vec{b})$ depends only on those items that are differently ordered by A and B.
6. The minimum positive distance is one.

The distance $d(\vec{a},\vec{b})$ is based on a comparison of the ordering of items i, j for A and B for each of the $\frac{1}{2}n(n-1)$ pairs (i, j). For any decision-maker A the preference order of n items can be summarized in a square $n \times n$ matrix A with elements a_{ij} $(1 < i,j < n)$

a_{ij} = +1 if (row-)element i is preferred to (column-) element j;

a_{ij} = -1 if (column) element j is preferred to (row-) element i;

a_{ij} = 0 if the preferences for i and j are tied, i.e. if they are preferred equally.

The diagonal elements of the matrix A are zero.

The unique distance between A and B, satisfying the axioms is:

$$d(\vec{a},\vec{b}) = \sum_{i<j} | a_{ij} - b_{ij} | = \sum_{i<j} d(a_{ij}, b_{ij}) \tag{4.13}$$

in which $d(a_{ij}, b_{ij}) = 0$ if items i and j are ordered in the same way,
$d(a_{ij}, b_{ij}) = 2$ if items i and j are ordered in the opposite way,
and $d(a_{ij}, b_{ij}) = 1$ if either A or B tied the items i and j.

The distance of all pairs of decision-makers can be computed by (4.13) and represented in a distance matrix (see also (4.2).

$$D = \| d(\vec{a},\vec{b}) \| \text{ of order } m \times m \tag{4.14}$$

Hazewindus and Mokken (1972) illustrate the procedure on some examples in terms of Daalder's data concerning the twelve parties in the Second Chamber of the Dutch Parliament at the time of his survey (1968). In the example of table 4.6 the distance of MP's A and B is minimal and equal to 1, because only their ordering of the two extreme Calvinist parties SGP and GPV differs; between A and C maximal and equal to $n(n-1)$, because they order the twelve parties oppositely.

	CPN	PSP	PvdA	D'66	PPR	KVP	ARP	CHU	VVD	BP	SGP	GPV
MP A	1	2	3	4	5	6	7	8	9	10	11	12
MP B	1	2	3	4	5	6	7	8	9	10	11,5	11,5
MP C	12	11	10	9	8	7	6	5	4	3	2	1

Table 4.6.

Rank orders of twelve parties by three hypothetical KVP MP's.

From (4.13) and the above examples we see that all differences in ordering have equal weight in the calculation of the distance: differences of rank ordering of major parties contribute to the distance as much as those of minor parties; differences among highest preferred items as much as among lowest preferred ones. This may be illustrated by another hypothetical example for three MP's of the former Catholic party KVP, one of the religious parties in the center, member of all post-war coalitions. Their preference vectors are given in tabel 4.7a. The three rank order differ only in their ordering of four parties: PvdA, VVD, SGP and GPV. By axiom 5 we need only consider the relevant submatrices (table 4.7b) to compute their distances (table 4.7c).

a. Rank orders.

	CPN	PSP	PvdA	D'66	PPR	KVP	ARP	CHU	VVD	BP	SGP	GPV
MP E	9	8	5	7	6	1	2	3	4	12	11	10
MP F	9	8	5	7	6	1	2	3	4	12	10	11
MP G	9	8	4	7	6	1	2	3	5	12	11	10

b. Submatrices of orderings between parties for which their ordering differed.

	E			F			G		
	PvdA	VVD	SGP	PvdA	VVD	SGP	PvdA	VVD	SGP
VVD	+1			+1			-1		
SGP	-1	-1		-1	-1		-1	-1	
GPV	-1	-1	+1	-1	-1	-1	-1	-1	+1

c. Distance matrix between the three hypothetical KVP MP's.

	E	F	G
E	0	2	2
F	2	0	4
G	2	4	0

Table 4.7.

Rank orders of twelve parties by three hypothetical KVP MP's, their relevant submatrices and distance matrix.

The distance between E and F is 2, because of the different ordering of the minor parties SGP and GPV. This distance is equal to the distance between E and G, which is caused by their highly important different ordering of the two major parties and coalition candidates, the socialist party PvdA and the conservative/liberal party VVD. Yet intuition will lead us to consider E and F much closer to each other than E and G. In fact we are faced here with a version of the important problem of intensity in the interpersonal comparison of ordinal utility.

4.2.2 *Normalized distance.*

The range of distances, as defined above, will depend on the number of items that are ordered. We can easily see that the maximum possible distance for a set of n items is:

$$d_{max} = n(n-1)$$

For purposes of comparison, more or less independent of the number of items we may prefer a normalized distance, as defined by

$$d^*(\vec{a},\vec{b}) = \frac{d(\vec{a},\vec{b})}{d_{max}} = \frac{d(\vec{a},\vec{b})}{n(n-1)} \tag{4.15}$$

Obviously, we have

$$0 \leq d^*(\vec{a},\vec{b}) \leq 1$$

for any set of items.

Corresponding to d^* and (4.14) we define the normalized distance matrix

$$D^* = \| d^*(\vec{a},\vec{b}) \| \tag{4.16}$$

It can easily be shown dat $d(\vec{a},\vec{b})$ and $d^*(\vec{a},\vec{b})$ are simple linear transforformations of Kendall's coefficient of rank correlation τ_{AB} between the rank orders $\vec{a}$ and $\vec{b}$:

$$\tau_{AB} = 1 - \frac{d(\vec{a},\vec{b})}{\frac{1}{2}n(n-1)} = 1 - 2\frac{d(\vec{a},\vec{b})}{d_{max}} = 1 - 2d^*(\vec{a},\vec{b}) \quad (4.17)$$

τ_{AB} may therefore be considered as an <u>index of agreement</u> between A and B, whereas $d^*(\vec{a},\vec{b})$ and $d(\vec{a},\vec{b})$ are measures of distance or disagreement.

4.2.3. *Reduced distance.*

As in section 4.1.3 we will relate the distance between decision-makers on the basis of their actual preference orders to the expected distance in the case of random ordering. The derivations of appropriate values of statistics of $d(\vec{a},\vec{b})$ under the assumption of random rank ordering follows straightforward from the distribution theory given by Kendall for τ (Kendall, 1955). It can be demonstrated that:

$$\mathcal{E}\{\underline{d}(\vec{a},\vec{b})\} = \tfrac{1}{2}n(n-1) \quad (4.18)$$

in case of random preference ordering. This results in the following <u>reduced distance</u>:

$$d^r(\vec{a},\vec{b}) = \frac{d(\vec{a},\vec{b})}{\frac{1}{2}n(n-1)} = \tfrac{1}{2}d^*(\vec{a},\vec{b}) \quad (4.19)$$

As we see here, in case of preference orderings, the reduced distance is a simple scalar transformation of the normalized distance because both the maximum and the expected distance depend only on the number of items n. This implies that the difference between normalized distances and reduced distances is less important in case of preference orders. As our measures and procedures of the next sections are equally applicable for roll call situations as well as for preference orders and as they are defined in terms of the reduced distance matrix, we introduce also <u>in case of preference</u> orders the <u>reduced distance matrix</u>

$$D^r = \| d^r(\vec{a},\vec{b}) \| \quad (4.20)$$

as the basis of our coefficients.

4.3. *Group analysis*

In this section a number of measures are proposed for the analysis of <u>a priori</u> defined groups of decision-makers. Group denotes any set of decision-makers that has been a priori defined independently of the observed voting behavior or preference orders.

First we want to characterize the different groups according to their cohesion in voting or preference ordering. Second, we want to evaluate the policy locations of decision-makers within the groups to which they belong in order to determine the degree to which a decision-maker contributed to the cohesion of the group as a whole. These are the two main perspectives of intra-group analysis. Moreover, we want to compare policy locations of different groups relative to each other: how are two or more groups located with respect to each other and what are the policy locations of the decision-makers in one group with respect to those in the other group? These are the main perspectives of inter-group analysis. Thus, we distinguish in group analysis four different perspectives on the basis of two dichotomies (1) measures characterizing groups as a whole versus measures indicating policy locations of decision-makers wich respect to the group; (2) intra-group analysis versus inter-group analysis. These perspectives are summarized in the following scheme:

	intra-group analysis	inter-group analysis
characteristics of groups as a whole	measures of group cohesion	measures of location between groups as a whole
policy locations of delegations relative to the group	member coefficients of location within group	member coefficients of location relative to the other group

All our measures are defined on the basis of the matrix of reduced distances D^r, which contains the reduced distances. It does not matter whether the matrix of reduced distances is based on roll calls (see expression (4.6)) or on preference orders (see expression (4.20)): all measures can be used in both situations and are defined in exactly the same way. We will therefore not distinguish these two situations anymore during the introduction of the measures. During the introduction all measures will be illustrated with examples of group analysis in the United Nations based on roll calls. After the introduction of the measures two more elaborate examples will be given, one for roll call analysis in the United Nations (section 4.3.4), and one for analysis of preference orders of Dutch MP's (section 4.3.5).

The matrix D^r of reduced distances is a square matrix; its rows and columns represent decision-makers. If we want to analyze two groups of decision-makers (e.g. in the United Nations General Assembly the Latin American and Afro-Asian groups), we order the rows and columns of the distance matrix

intra-group analysis Latin-America	inter-group analysis Latin America-Afro Asia
inter-group analysis Latin America-Afro-Asia	intra-group analysis Afro-Asia

Figure 4.1

Intra-group and inter-group analysis of the Latin-American group and the Afro-Asian group. D^r: matrix of reduced distances between member states in the UN.

according to these groups. The interpair distances between members of the same group then form a diagonal block in D^r (see figure 4.1). Measures of inter-group analysis are based on the off-diagonal rectangles of the matrix, that contain the interpair distances of the decision-makers in the first group (Latin America) relative to the decision-makers in the second group (Afro-Asia) (see figure 4.1). If two groups partially overlap, the inter-group analysis will be based only on the distances between the non-overlapping decision-makers. For example, an inter-group analysis between the partially overlapping Asian and Arab groups will be based only on the distances between the African-Arab delegations and the non-Arab-Asian delegations.

Measures that characterize groups as a whole are based on the distances in the whole submatrix within or between the groups. Measures that indicate policy locations of decision-makers relative to the group are based on the row or vector of distances of that decision-maker with the (other) group members.

We shall treat in section 4.3.1 the coefficients and measures that we use to characterize groups as a whole, and in section 4.3.2 the indicators of the policy locations of the decision-makers to the group. In both sections we shall only consider intra-group analysis. In section 4.3.3 we shall consider how these measures should be redefined for application in the inter-group analysis (for the analysis of the off-diagonal rectangles of the distance matrix).

We suppose that we analyze a set L of m decision-makers. A group is a subset $G \subset L$ of k decision-makers ($k \leq m$). We designate by $g_1, g_2, \ldots, g_k$ the decision-makers belonging to $G(g_i \in G)$ and their corresponding vote or preference vectors by $\vec{g}_1, \vec{g}_2, \ldots \vec{g}_k$. The intra-group analysis will be based on the $k \times k$ reduced distance matrix $D^r(G)$ for the k group members. In table 4.8 we give the distances for the Commonwealth group over roll calls on colonial and socio-economic issues in the period 1950-1955 (see Stokman, 1977). Equal probabilities for the different voting alternatives are chosen. All measures of intra-group analysis will be illustrated on the basis of this distance matrix.

	New Zealand	Canada	Australia	United Kingdom	South Africa	India	Pakistan	sum distance (sd_i)	average distance $a(g_i) = \frac{1}{k-1} sd_i$	coefficient of location $c_4(g_i) = 1 - a(g_i)$
New Zealand	-	.21	.21	.24	.39	1.37	1.47	3.89	.65	.35 pivot
Canada	.21	-	.27	.28	.43	1.34	1.43	3.96	.66	.34
Australia	.21	.27	-	.17	.31	1.52	1.59	4.07	.68	.32
United Kingdom	.24	.28	.17	-	.33	1.50	1.58	4.10	.68	.32
South Africa	.39	.43	.31	.33	-	1.53	1.59	4.58	.76	.24
India	1.37	1.34	1.52	1.50	1.53	-	.33	7.59	1.27	.27
Pakistan	1.47	1.43	1.59	1.58	1.59	.33	-	7.99	1.33	.33
								18.09	.86	.14

range of the group $r(G)$

coefficient of group cohesion $c_1 = 1 - r(G) = -.59$

range pairs: Pakistan - South Africa
Pakistan - Australia

total sum-distance in the group $sd = \frac{1}{2}\Sigma\, sd_i$

average distance in the group $a(G) = \frac{2}{k(k-1)} sd$

coefficient of group cohesion $c_4 = 1 - a(G)$

Source: Stokman, 1977, 96.

Table 4.8.

Matrix of reduced distances and measures of intra-group analysis for the Commonwealth group over the selected roll calls on colonial and socio-economic issues in the period 1950-55; equal probabilities; absence considered as abstention.

4.3.1. Intra-group analysis: characteristics of the group as a whole.

A major variable used to characterize a group of decision-makers is the group cohesion. With perfect group cohesion all members take the same policy positions, resulting in distances of zero between all group members. In this case:

$$D^r(G) = 0.$$

In general, however, we observe differences in votes or preferences and therefore positive elements in $D^r(G)$. Several measures of cohesion may be defined. One measure is to consider the most divergent policy positions in a group. A group will be less cohesive, if it contains a broad range of policy positions. The range of policy positions in a group is reflected in the value of the maximum interpair distance in the matrix $D^r(G)$. It gives the maximum distance to be found within the group. Our first measure of cohesion is based on this range. We define, therefore, as our first coefficient of group cohesion ($c_1(G)$):

$$c_1(G) = 1 - r(G) \qquad (4.21)$$

The maximum value of $c_1(G)$ is 1, if and only if all group members voted identically on all n roll calls or had identical preference orders. It should be noted that $c_1(G)$ can assume negative values for cases where a pair of group members has a larger distance than expected from random voting behavior or preference ranking. Obviously, we should have for cohesive groups:

$$0 < c_1(G) \leq 1.$$

In table 4.8 we see that the range of the Commonwealth group was 1.59 in the period 1950-1955. It is observed between Pakistan and Australia as well as between Pakistan and South Africa. The resulting strongly negative value of the coefficient of cohesion $c_1(c_1 = -.59)$ indicated that the Commonwealth group cannot be considered as cohesive.

One may question the usefullness of $c_1(G)$ as a measure of cohesion, because it is based on only one distance in the group: it is completely determined by the extreme positions in the group. For instance, it may be possible that a group in which all members except one always voted together (c.q. had the same preference orders) and another group in which all members were voting relatively heterogeneously (c.q. having different preference orders) will each have the same value of c_1. Moreover, the distribution theory of

c_1 is difficult to derive. For these reasons it is desirable to introduce another measure of cohesion, one that is based on all distances within the group. In order to take all distances within the group into consideration for the determination of the group cohesion, we define a second measure of cohesion on the average interpair distance for a group G. The average interpair reduced distance of a group G, $a(G)$, is given by:

$$a(G) = \frac{2}{k(k-1)} \Sigma_{i<j}\, d^r(g_i,g_j) = \frac{2}{k(k-1)} sd \qquad (4.22)$$

in which sd denotes the sum of the $\frac{1}{2}k(k-1)$ distances in the group G. $a(G)$ can now be used to define a coefficient of group cohesion that is based on all distances within the group. Following Mokken and Stokman (1970, 4-6) we denote this coefficient by $c_4(G)$:[8]

$$c_4(G) = 1 - a(G) \qquad (4.23)$$

Again, as in the case of $c_1(G)$, this coefficient $c_4(G)$ has the maximum value of 1, if and only if all group members voted identically or had identical preference orders. The coefficient is 0, if the average reduced distance in the group is equal to 1. In that case the average absolute distance in the group is equal to that expected in a group of randomly and independently voting or preference ranking decision-makers. For a cohesive group we should have[9]

$$0 < c_4(G) \leq 1$$

The coefficient $c_4(G)$ can also reach negative values. This occurs when the average reduced distance in the group is larger than 1, in other words when the average absolute distance in the group is larger than that expected in a group of randomly and independently voting or preference ranking members. For a group to be incoherent we should have:[10]

$$c_4(G) < 0$$

In table 4.8 we see that the average distance in the Commonwealth group was .86. In terms of its coefficient of cohesion c_4 the Commonwealth group can therefore be considered as cohesive in the period 1950-1955 (c_4 = .14). This positive value is due to the fact that only two group members took deviant policy positions, namely India and Pakistan. The rest of the group votes rather simularly. The positive value of $c_4(G)$ and the negative value of $c_1(G)$ in table 4.8 illustrate the different meanings of the two coefficients of cohesion.

The coefficient $c_4(G)$ of group cohesion has a number of desirable properties. Its theoretical maximum (1) and its theoretical minimum (0) are invariant for different numbers of roll calls (n) or items in preference orders and different numbers of group members (k). This makes it very useful in research designs, in which the cohesion of groups of very different size over different periods of analysis are analyzed. Our coefficient also has the desirable property of enabling the individual contribution of the group members to the group cohesion to be evaluated in terms of it. In other words, the coefficient will differentiate marginal group members from central group members, as we shall see in the next section. These advantages of $c_4(G)$ stand in sharp contrast to some other proposed coefficients of cohesion. For instance, Willetts (1972, 573; 1978) who used Lijphart's index of agreement (Lijphart, 1963), needs three indices of agreement for a legislature or a group of legislations:

$$I_\alpha = \frac{\text{Number of pairs of members in significant agreement}}{\text{total possible number of pairs of members}}$$

$$I_\beta = \frac{\text{Number of pairs of members in significant disagreement}}{\text{total possible number of pairs of members}}$$

$$I_\gamma = I_\alpha - I_\beta$$

He states:

"In theory it will be possible for I_γ to obtain a negative value by the number of pairs in disagreement being higher than the number of pairs in agreement, but this is not very likely to occur in practice. I_γ should not be reported without the corresponding values of I_α and I_β and all three should be given at the chosen level of significance. This is because both I_α and I_β may become higher at a lower level of significance." (Willetts, 1972, 573-4).

To compare different groups in terms of those indices would be very difficult, because three values have to be taken into consideration. Moreover, it would be difficult to evaluate the cohesion of a group over different issues and/or time, because at the same level of significance the test of significance will gain power if the number of issues is larger. Finally, groups of different size can not be compared on the basis of these coefficents because the covariance between the interpair indices of agreement within a group has not been taken into consideration. In section 4.1.3 we stated that the maximum expected distance in case of random voting is obtained if the probability of voting in favor and against is equal to .5 ($\pi_1 = \pi_{-1} = \frac{1}{2}$; $\pi_0 = 0$). Using the reduced distances based on these probabilities, we are now able to judge the level of cohesion of a whole decision-making body, such as the UN General Assembly, as measured against this criterion of maximal expected distances. As an example in table 4.9 the coefficients of group cohesion are given for the United Nations General Assembly as a whole for the four periods and over the roll calls on colonial and socio-economic issues that were the basis of the analysis of Third World Group formation in the study of Stokman (1977). In the coefficient of

	Period			
	1950-55	1956-59	1960-63	1965-68
c_4	.24	.28	.36	.42
c_1	-.59	-.47	-.35	-.47
range pairs	Belgium-Saudi Arabia	United Kingdom-Poland	Belgium-Ukr.SSR Belgium-USSR	Portugal-Syria

Source: Stokman, 1977,101

Table 4.9.

Coefficients of group cohesion for the UN General Assembly as a whole over roll calls on colonial and socio-economic issues per period of analysis; probabilities in favor and against .5; absence considered as abstention.

c_4 the growing consensus on these issues in the UN General Assembly is reflected by a steadily increasing value from .24 in the first period to .42 in the last. The coefficient of cohesion c_1 increased also in the first three periods, but declined again in the last period. The table illustrates once again the different meanings of the two coefficients. The support of a vast majority in the UN for decolonization caused a decreasing average distance, whereas a small resistant minority caused an increasing maximum distance.

4.3.2 Intra-group analysis: location of members in the group

The reduced distance between two decision-makers gives us information about their relative policy positions. However, we want to measure the policy locations of group members relative to the group as a whole, i.e. relative to all other group members. In such a measure all reduced distances between that decision-maker and all other decision-makers can best be taken into account. To achieve this, we consider the average reduced distance of group member g_i, $a(g_i)$ to the other $k-1$ group members. This average reduced distance is given by:

$$a(g_i) = \frac{1}{k-1} \sum_{\substack{j=1 \\ j \neq i}}^{k} d^r(g_i, g_j) = \frac{1}{k-1} sd_i \qquad (4.24)$$

in which sd_i denotes the sum of the reduced distances of g_i with the other $k-1$ group members.

Just as $a(G)$ was used in the last section to define the coefficient of group cohesion $c_4(G)$, $a(g_i)$ can now be used to define a coefficient of location within the group. Because this coefficient can be considered as the item coefficient of $c_4(G)$, we shall denote that coefficient of location of g_i within the group by $c_4(g_i)$ or $c_4^{(i)}$. It is defined by

$$c_4(g_i) = 1 - a(g_i) \qquad (4.25)$$

The maximum value of $c_4(g_i)$ is 1; in that case all (reduced) distances are zero. The coefficient of location is 0, if the average reduced distance is 1; in that case the average absolute distance of that member is equal to that expected in the case of randomly and independently voting or preference ranking group members. For a 'loyal' group members we should have:[11)]

$$0 < c_4(g_i) \leq 1$$

The coefficient $c_4(g_i)$ can be used to <u>order the group members from central to marginal in the group</u> by permutation in such a way that:

$$c_4(g_1) \geq c_4(g_2) \geq \ldots \geq c_4(g_i) \geq \ldots \geq c_4(g_k)$$

This quasi-order can be used to define the <u>pivotal</u> or 'central' members of a group. <u>A pivotal member of a group G</u>, denoted by g_p, is a member of which $c_4(g_i)$ is maximal. That member has the smallest sumdistance (average distance) relative to the other group members. Sometimes more than one group member has this coefficient of location.

In table 4.8 the sumdistance, average distance, and coefficient of location are given for each of the members of the Commonwealth group. The group members are ordered from central to marginal. New Zealand had the highest coefficient of location; it was the pivot of the group. India and Pakistan took very marginal positions in the group, having strongly negative coefficients of location. These locations in the group changed drastically after 1955 as a result of the entrance of new group members. This is shown in table 4.10, in which we compare the Commonwealth group for two periods of analysis, 1950-1955 and 1960-1963. In 1960-1963 the positions of the old Commonwealth states (New Zealand, Canada, Australia, United Kingdom) became marginal, whereas those of India and Pakistan became central. The cohesion of the group increased considerably.

(a) Period 1950-55	$c_4(g_i)$	(b) Period 1960-63	$c_4(g_i)$
New Zealand	.35	Malaysia	.46
Canada	.34	Cyprus	.43
Australia	.32	Pakistan	.41
United Kingdom	.32	Nigeria	.40
South Africa	.24	India	.40
India	-.27	Ceylon	.36
Pakistan	-.33	Ghana	.32
		Canada	.29
		New Zealand	.28
		Australia	.22
		United Kingdom	.12
coefficient of group cohesion c_4 = .14		coefficient of group cohesion c_4 = .34	
coefficient of group cohesion c_1 = -.59		coefficient of group cohesion c_1 = -.35	
range pairs: Pakistan-Australia, Pakistan-South Africa		range pair: United Kingdom-Ghana	

Source: Stokman, 1977, 104

Table 4.10.

Coefficients of location and group cohesion of the Commonwealth group over roll calls on colonial and socio-economic issues in the periods 1950-55 and 1960-63; equal probabilities; absence considered as abstention.

The theoretical significance and implications of the concept of pivotal member are not yet clear. The fact that the pivotal member of a group is the one with the minimum average distance to the other group members seems to imply an intermediate position between the other group members in the policy area covered by the given set of votes. As such, the pivotal member, in the multidimensional case we are investigating, may bear some resemblance to the median member introduced by Black (1963, 18) in his theory of majority voting in committees on one dimensional policy areas. It may be worthwile to investigate further whether the pivotal member may be thought to indicate a central position analogous to the median position that has proven to be optimal in the determination of group decision under the special assumptions of Black's theory. In that case, the pivotal member may be used to characterize the general location of the group to which it belongs. As such, the position of the pivotal member can be considered as a characteristic of the group as a whole as well as an indicator of the position of group members within the group. Thus, it has two functions: it indicates the central group members, but also indicates the policy locations of the group as a whole.

The ordering of the members on the basis of the coefficient of location does not tell us much about possible coherent blocs of group members within a group. If we are interested in the different factions within a group, the procedures of bloc- and clique analysis, given in section 4.4 and 4.5 can be used.

4.3.3 *Inter-group analysis*

In this section we consider two groups of decision-makers: group G_1 with k group members $g_1,\ldots,g_k$ and group G_2 with t group members $g_1,\ldots,g_t$. We suppose that these groups as disjoint. The analysis of partially overlapping groups will only be based on the nonoverlapping members. As an illustration the reduced distances between the members of the Commonwealth group and those of the Scandinavian group over roll calls on colonial and socio-economic issues for the period 1950-55 from the Stokman 1977 study are given in table 4.11. All measures of inter-group analysis will be illustrated on the basis of this distance matrix.

In inter-group analysis our first objective is to locate the groups as a whole relative to each other. For this purpose we introduce four measures. First, the use of pivotal members of groups as indicators of the location of groups serves to define distances between groups. We define the distance between two groups G_1 and G_2 as the minimal reduced distance between their respective pivotal members. Thus, the distance between the Common-

	Norway	Denmark	Sweden	Iceland	with respect to Scandinavian group: sum-distance sd_i	average distance $a(g_i,G_2)$	coefficient of location $c_4(g_i,G_2)$
New Zealand	.39	.41	.38	.77	1.95	.49	.51
Canada	.37	.42	.38	.77	1.94	.49	.51
Australia	.48	.49	.50	.80	2.27	.57	.43
United Kingdom	.50	.50	.51	.82	2.33	.58	.42
South Africa	.60	.63	.59	.73	2.55	.64	.36
India	1.15	1.18	1.19	1.06	4.58	1.15	-.15
Pakistan	1.27	1.27	1.22	1.06	4.82	1.21	-.21
With respect to Commonwealth group:							
sumdistance sd_j	4.76	4.90	4.77	6.01	20.44		
average distance $a(g_j,G_1)$	.68	.70	.68	.86		.73	
coefficient of location $c_4(g_j,G_1)$	.32	.30	.32	.14			.27

distance between the most proximate pair

inter-pivot distance

inter-group range $r(G_1,G_2)$

inter-group coefficient of cohesion $c_1(G_1,G_2) = 1 - r(G_1,G_2) = -.27$

total sumdistance between groups $sd = \Sigma sd_i = \Sigma sd_j$

average distance between groups $a(G_1,G_2)$

intergroup coefficient of cohesion $c_4(G_1,G_2)$ $1 - a(G_1,G_2)$

Source: Stokman, 1977, 106-7

Table 4.11

Matrix of reduced distances and measures of inter-group analysis between the Commonwealth group and the Scandinavian group over roll calls on colonial and socio-economic issues in the periode 1950-55; equal probabilities; absence considered as abstention.

wealth group and the Scandinavian group was .39 (see table 4.11). Second, as in intra-group analysis, the range of policy positions between the two groups is a relevant fact. We define therefore the inter-group range $r(G_1, G_2)$ between two groups G_1 and G_2 as the maximal reduced distance in the off-diagonal part of the distance matrix. The intergroup coefficient of cohesion $c_1(G_1,G_2)$ is based on this range and is defined by:

$$c_1(G_1,G_2) = 1 - r(G_1,G_2) \tag{4.26}$$

The inter-group coefficient of cohesion $c_1(G_1,G_2)$ between the Commonwealth and the Scandinavian group was -.27 (see table 4.11). Here we see that the cohesion between two groups might well be higher than that within one of the groups: the inter-group coefficient of cohesion is higher than the coefficient of cohesion within the Commonwealth group itself (see table 4.8).

A third measure to locate the groups relative to each other is the reduced distance between the most proximate pair: the minimal observed distance between a member of group G_1 and a member of group G_2. The most proximate pair is not necessarily unique. Canada and Norway formed the most proximate pair between the Commonwealth and Scandinavian group. Their distance (.37) was only somewhat smaller than that between the two groups (see table 4.11).

The fourth and final measure for the location of the groups to each other is a coefficient of cohesion c_4 adapted for inter-group analysis. It is based on the average reduced distance in the off-diagonal part of the matrix, containing the reduced distances between members of the two groups G_1 and G_2. This average reduced distance between the groups G_1 and G_2 - denoted by $a(G_1,G_2)$ - can be used to define the intergroups coefficient of cohesion $c_4(G_1,G_2)$ between two groups G_1 and G_2:

$$c_4(G_1,G_2) = 1 - a(G_1,G_2) \tag{4.27}$$

The maximum values of $c_4(G_1,G_2)$ is 1; in that case all (reduced) distances between the two groups are zero. The coefficient is 0, if the average reduced distance is 1; in that case the average absolute distance between the two groups is equal to that expected in case of random and independent voting or preference ranking.[12)] Table 4.11 shows that the inter-group coefficient of cohesion $c_4(G_1,G_2)$ between the Commonwealth and the Scandinavian group was .27.

A second important objective in inter-group analysis is to indicate the policy locations of individual group members relative to the other group as a whole: which decision-makers have a close policy location and which have a distance policy location relative to the other group? For this purpose we define a coefficient of location relative to the other group that is very similar to the coefficient of location within the group (see section 4.3.2). It is based on the average reduced distance of the group member g_i ($g_i \in G_1$) with respect to the group members in group G_2. If we denote this average reduced distance by $a(g_i, G_2)$, the <u>coefficient of location of g_i ($g_i \in G_1$) with respect to group G_2</u> is defined as follows:

$$c_4(g_i, G_2) = 1 - a(g_i, G_2) \tag{4.28}$$

The maximum value of $c_4(g_i, G_2)$ is 1; the coefficient is 0, if the average absolute distance from g_i to the group members of G_2 is equal to that expected in the case of random and independent voting or preference ranking. 13) On the basis of this coefficient we can order the group members of group G_1 from close to distant with respect to group G_2. In table 4.11 New Zealand and Canada were closest to the Scandinavian group; their coefficient of location was .51. India and Pakistan were most distant from the Scandinavian group with coefficients of location of -.15 and -.21. Norway and Sweden were closest to the Commonwealth group (coefficient of location: .32), followed by Denmark (.30) and Iceland (.14).

All measures in the inter-group analysis are defined on the k by t off-diagonal matrix, in which the reduced distances between the two groups are given. Of course, it is also possible to analyze the union of the groups G_1 and G_2 with the measures developed for intra-group analysis in the last two sections. This results in pivot(s), coefficients of location within and coefficients of cohesion of the overall group ($G_1 \cup G_2$). (In our example of figure 4.1 it results in the analysis of the whole group of developing nations). In the next two sections we will illustrate the usefulness of this apporach on the basis of two more elaborated illustrations: one on Third World groups in the UN General Assembly on the basis of roll call analysis and one on Dutch political parties and possible coalitions on the basis of preference orders of twelve Dutch parties by Dutch MP's.

4.3.4 Illustration 1: Voting cohesion of groups of developing nations

In this illustration, based on the study of Stokman (1977), policy locations and voting cohesion of groups of developing nations are examined over four periods of analysis on the basis of roll calls on colonial and socio-economic issues. Different groups of developing nations are considered. The Latin-American, Asian and African groups are three disjoint groups of developing nations. The Afro-Asian group is the union of the Asian and African groups plus Yugoslavia and Cyprus; the group of all developing nations is the Union of the Latin-American and Afro-Asian group. Moreover, the Arab group is considered. This group partially overlaps the Asian and African group. Particularly the Asian, African, Afro-Asian and Arab groups have shown a large increase of membership over the period of analysis.

Over time we can perceive at least two opposing forces that could affect the voting cohesion of groups of developing nations: on the one hand, we expect a larger group cohesion among the developing nations over time because of a growing groups consciousness; on the other hand, the maintenance of a high level of group cohesion will be more difficult in later periods, because of the larger group size. A high voting cohesion is a primary group function of the developing nations: developing nations in the General Assembly must be present and vote alike in order to support their own proposals and to bloc undesired alternatives; if not, their majority position cannot be effective.

Table 4.12 contains the coefficients of cohesion c_4 for the different groups of developing nations over the four periods of analysis. Absense is treated as abstention. Absence is and indicates nonparticipation in the decision-making; from this point of view, absence weakens the cohesion of a group as an instrument of influence and may not be considered as missing data. In table 4.12 we used for all four periods the same probabilities for the different voting alternatives ($\pi_1 = \pi_0 = \pi_{-1} = 1/3$). This implies that we compare the voting cohesion of the groups over the four periods, without taking into consideration the voting cohesion of the UN General Assembly as a whole: table 4.12 therefore gives information about the absolute differences in voting cohesion of a group over the four periods of analysis.

	1950-55	1956-59	1960-63	1965-68
Latin America	.36	.37	.49	.52
Asia	.63	.50	.53	.51
Africa	.58	.70	.49	.64
Arab	.74	.71	.67	.71
Afro-Asia	.62	.54	.49	.58
developing nations	.38	.30	.40	.50

Source: Stokman, 1977, 169.

Table 4.12.

Voting cohesion of groups over all selected roll calls per period; absence = abstention; equal probabilities.

The group of developing nations became more cohesive regarding colonial and socio-economic issues from the second to the fourth period (.30, .40, .50); in the first period the developing nations had about the same cohesion as in the third period, namely .38. The higher cohesion in the later periods was primarily caused by the higher cohesion among the Latin-American countries (.36, .37, .49, .52) and, as we shall see later, the smaller distances between the Latin-American countries and the Afro-Asian countries. The Afro-Asian group was least cohesive in the third period (1960-63): .49.

That low level of cohesion was caused by differences in policy positions among the African delegations: for the African group the level of cohesion was also lowest is the third period (.49). It shows that the African controversies, which resulted in two different groups of African states, the Casablanca and the Brazzaville groups, had farreaching consequences for the voting cohesion of the whole Afro-Asian group. The Arab group was the most cohesive group in all four periods of analysis; its cohesion was remarkably stable over time, unaffected by the different Arab controversies outside the United Nations. The Asian group had a high cohesion in the first period, reflecting a great unity among the Asian countries around the time of the Bandung Conference, when the Afro-Asian initiative was primarily located in the Far East.

In table 4.9 we saw that the absolute level of cohesion of the UN General Assembly as a whole became higher over time. The coefficient c_4 grew from

.24 in the first period to .42 in the last period. Groups that have the same level of cohesion over the four periods in table 4.12 maintained their level of cohesion within a more cohesive General Assembly; their relative level of cohesion became lower over time. If the coefficients of cohesion c_4 are based on reduced distances with estimated probabilities for the voting alternatives we observed indeed a lower level of cohesion in the later periods for the Asian, Arab and Afro-Asian countries. The cohesion of the developing nations in the fourth period, as measured against this standard, reached the same level as that of the first period. The four coefficients of cohesion c_4 of the developing nations were then: .36, .27, .33 and .37.

From table 4.12 it becomes clear that the developments in the cohesion of the group of developing nations cannot be explained by developments in cohesion of the Afro-Asian group. The relations between the Latin-American countries and the Afro-Asian countries have therefore special importance. In tabel 4.13 we give the results of the inter-group analysis of the Latin-American group with the different groups of Afro-Asian countries.

	1950-55	1956-59	1960-63	1965-68
a) inter-group coefficients of cohesion c_4 of Latin America with:				
Asia	.30	.16	.30	.39
Africa	.33	.08	.29	.38
Arab	.30	.06	.23	.36
Afro-Asia	.30	.13	.29	.39
b) inter-pivot distances of Latin America with:				
Asia	.61	.79	.76	.59
Africa	.55	1.02	.60	.63
Arab	.61	1.02	.79	.55
Afro-Asia	.60	.79	.70	.63

Table 4.13

Inter-group coefficients of cohesion c_4 and inter-pivot distances of the Latin-American group with the groups of Afro-Asian countries per period; absence = abstention; equal probabilities.

Table 4.13a shows that the inter-group coefficients of cohesion between Latin-America and Afro-Asia gradually increased from .13 in the period 1956-59 to .39 in the period 1965-68. The distances between Latin-American countries and Afro-Asian countries were therefore remarkably smaller in the period 1965-68 than in the period 1956-59. This also holds for the distances between the pivots of the two groups: in table 4.13b it is shown that the inter-pivot distance between Latin-America and Afro-Asia decreased from .79 in 1956-59 to .63 in 1965-68. The same trend can be observed in table 4.13 for the different subgroups of Afro-Asian countries. There is only one exception: the inter-pivot distance between Latin-America and Africa in the third period was somewhat smaller than in the fourth period, due to the moderate, pro-western policy positions of the Brazzaville group in the period 1960-63. In the first period the distances between the Latin-American countries and the Afro-Asian countries were relatively small, all coefficients for the first periods having a level somewhere between those of the third and fourth period. <u>From table 4.13 we conclude that the shifts in cohesion of the group of developing nations as a whole were primarily determined by the relations between the Latin-American countries and the Afro-Asian countries</u>.

The smaller distances between Latin-America and Afro-Asia over the last three periods confirm the generally observed process of reorientation of Latin-America towards Afro-Asia. It indicates that the formation of the 'Group of 77' in 1963 was simply a manifestation of this larger process, rather than an isolated event. The high voting cohesion between Latin America and Afro-Asia in the first period was unexpected, however. How can it be explained? It might be due to the fact that the relations between Latin-America and Afro-Asia were not yet impeded by different positions regarding the East-West controversy during the first years of the period 1950-55. Most Afro-Asian countries had traditionalist, wester-oriented regimes in the beginning of that period; only a few of the 17 Afro-Asian countries adhered to the principles of nonalignment. In the second half of the period Afro-Asian group formation and nonalignment in the East-West controversy became two sides of the same coin. From the western point of view, colonial and socio-economic issues became even more intertwined with the Cold War when in 1955 the Soviet Union decided to establish friendly relations with nonaligned regimes. In this situation Latin-Amnerican policy positions became more distant from those of the Afro-Asian countries, in particular after 1954, when United States' intervention in Guatemala ended the most neutralist regime in Latin-America at that time.

The 'Group 77' in which all developing nations coordinate their policies with respect to socio-economic questions in the Genral Assembly of the UN and UNCTAD, can be viewed as the bridge between the Latin-American and Afro-Asian groups. As such, it can be seen as a coalition of different subgroups of developing nations. For a subgroup the attractiveness of the group of developing nations as a coalition depends not only on the cohesion of the overall group (as given in table 4.12), but also on the distance between its policy locations and that of the overall group. In section 4.3.3 we proposed the minimum distance between pivots of groups as an indicator of the distance between the policy locations of the two groups. Table 4.14 gives these minimal inter-pivot distances of the group of developing nations with the different subgroups. It shows that in all periods the policy location of the overall group was far closer to that of the Afro-Asian group than to that of the Latin-American group. In the first two periods the pivot of the Afro-Asian group was even the same delegation as that of the overall group, resulting in an inter-pivot distance of zero. For the Latin-American group the inter-pivot distance with the group of developing nations had the same level in three periods (about .60) and was only higher in the second period (.79). For the fourth period this distance was higher

	1950-55	1956-59	1960-63	1965-68
inter-pivot distances of the group of developing nations with:				
Latin America	.60	.79	.60	.59
Asia	.12	.00	.38	.29
Africa	.38	.30	.35	.16
Arab	.12	.30	.39	.23
Afro-Asia	.00	.00	.30	.16

Table 4.14.
Minimal inter-pivot distances of the group of developing nations with the subgroups of developing nations per period; absence = abstention; equal probabilities.

than we would have expected from tables 4.12 and 4.13. In the first two periods of analysis the policy location of the overall group of developing nations was closer to that of the Asian group than to that of the African group; in the last period it was closer to that of the African group. This indicates that the center of the group of developing nations was always very close to the center of the Afro-Asian group, and it explains why Afro-Asian countries were the main initiators of the 'Group of 77'.

4.3.5 *Illustration 2: Location and cohesion of political parties and coalitions in the Dutch parliament on the basis of MP's preference orders.*

As a second illustration of the usefulness of our distance metric and measures of group analysis we report an analysis, on the following question in Daalder's 1968 survey of members of the Dutch Second Chamber (Daalder and Rusk, 1970, 29; Daalder and Van de Geer, 1977):[14)]

"In foreign studies a good deal of attention has been paid to the question of what distances exist among different parties. We have been asked to investigate this problem for the Netherlands and we would very much appreciate your help. Here are some cards. Each card has the name of a party which is now represented in Parliament. Would you please arrange these cards in such order that the party on top is the party to which you yourself feel closest, the next one the second-closest, and so on?"

Apart from his own party each responding MP ordered the other eleven parties. By assumption each respondent rated his own party first. Of a total of 150 members of the Second Chamber 141 MP's responded to the questionnaire. The eleven non-responding MP's contained all members of the Dutch communist party (CPN: 5 MP's), 4 belonged to the Peasant Party (BP), 1 to the SGP (the only member of that party of the time) and 1 to D'66. Two MP's did not respond to the rank-ordering question. The results reported here cover therefore a total of 139 MP's.

As an example of possible analyses that might be done, table 4.15 contains the absolute and reduced distances between party pivots. Three blocs of parties can easily be discerned:

1) PvdA (the Socialist Party), D'66 (Democratics '66) and PPR (Radicals) with a maximum distance of 7 (d^r = .11). These three progressive parties

a. Absolute distances

PvdA	13									
D'66	19	6								
PPR	18	5	7							
KVP	51	40	44	39						
ARP	49	40	42	35	12					
CHU	59	50	52	49	14	14				
VVD	55	44	42	46	26	28	22			
BP	107	102	108	101	68	66	56	70		
SGP	97	88	91	88	53	53	39	57	17	
GPV	91	82	76	77	48	42	36	50	23	15
	PSP	PvdA	D'66	PPR	KVP	ARP	CHU	VVD	BP	SGP

b. Reduced distances

PvdA	.20									
D'66	.29	.09								
PPR	.27	.08	.11							
KVP	.77	.61	.67	.59						
ARP	.74	.61	.64	.53	.18					
CHU	.89	.76	.79	.74	.21	.21				
VVD	.83	.67	.64	.70	.39	.42	.33			
BP	1.62	1.55	1.64	1.53	1.03	1.00	.85	1.06		
SGP	1.47	1.33	1.38	1.33	.80	.80	.59	.86	.26	
GPV	1.38	1.24	1.15	1.17	.73	.64	.55	.76	.35	.23
	PSP	PvdA	D'66	PPR	KVP	ARP	CHU	VVD	BP	SGP

Table 4.15

Distances between party pivots.

formed at the time of the survey an electoral alignment, called PAK ('het Progressief Accoord'). It can be seen that the fourth left wing party, the PSP (Pacifist Socialist Party) is closely related to this bloc.

2) The three Christian center parties: KVP (Catholic Party), ARP and CHU (protestant parties) with a largest distance of 14 (d^r = .21).

A few years later these three parties started a close cooperation in a federation of these parties, the CDA (Christian Democratic Appel), which finally resulted in a merger of the three parties.

3) A third, less coherent bloc of the three right wing parties BP (Peasants Party), SGP and GPV (two orthodox protestant parties). Only between these party pivots and pivots of other parties reduced distances larger than 1 can be observed, i.e. larger than expected in case of random preference ranking.

Daalder and Rusk (1972) who investigated also possible coalitions stated: "...that a majority of Antirevolutionaries and Catholics feel closer to the Socialists than to the Liberals even though they formed a coalition with the Liberals at the time of our survey".

The results of table 4.15 make a coalition between the three Christian center parties (KVP, ARP and CHU) with the Liberals (VVD) instead of with the Socialist Party (PvdA) less surprising.
The inter-pivot distances of all three Christians center parties are smaller with the Liberals (VVD) than with the Socialists:

	Socialist Party (PvdA)	Liberal Party (VVD)
KVP	.61	.39
ARP	.61	.42
CHU	.76	.33

These results are substantiated if we consider the distances between party pivots and pivots of possible coalitions. The pivot of the three Christian center parties together (CDA) is the same as the pivot of the Catholic Party. If we evaluate the two possible coalitions - CDA with the Liberals (VVD) and with the Socialists (PvdA) -, on the basis of their respective distances between the pivot of the coalition and that of the CDA, the government coalition that was actually formed (VVD + CDA) was the best one from the point of view of the CDA. Its divergence as measured by the distance between the pivot of the coalition and that of the CDA (d^r = .06) was much smaller than that of the alternative coalition (d^r = .18).

Moreover, also from the points of view of the Liberals and Socialist Party the existing government coalition was more attractive, because it entailed a much smaller shift in pivot for the Liberals (d^r = .39) than for the PvdA (d^r = .61). These values can be observed in Table 4.15, because the pivot of the CDA is the same as the pivot of the KVP.

Of course, above analyses might well be extended by analyses on the cohesion of parties and possible coalitions. As they are very similar to those given in the illustration of the last section on developing countries in the UN, we will not present them here.

4.4 *Bloc analysis*

In the analysis of a priori defined groups (section 4.3.1) we introduced two coefficients of group cohesion. The first coefficient $c_1(G)$ is based on the range of the distances in the group; the second coefficient c_4 (G) is based on the average distance in the group. We saw that the coefficients are related to two different aspects of group cohesion. In this and the next section we do not consider a priori defined groups of decision-makers. Instead of an analysis of a priori defined groups of decision-makers in terms of cohesion and location, we now want a procedure to define blocs or clusters of decision-makers in terms of close policy positions. Two different procedures are available for this purpose. The first, bloc analysis, is primarily based on the coefficient of cohesion c_4 and, by consequence, on the average distance in a group or bloc of decision-makers. The second, clique analysis, is based on the coefficient of cohesion c_1 and the range within a group or bloc of decision-makers (see section 4.5). The procedure of bloc analysis can be described as follows:

1. The smallest distance in the distance matrix is chosen to define a bloc of two decision-makers. The decision-maker that has a minimum average distance to the first two decision-makers is then added to the bloc. In this way the process is continued; in each step the bloc is extended to include the decision-maker that has a minimum average distance to those previously chosen. This process is repeated until either c_1 or c_4 falls below a certain minimum level q_1, q_4, the bloc-defining constants.
2. After the final determination of the first bloc, a second bloc is formed from the rest of the matrix in the same way. This process is repeated until no other blocs can be found that satisfy the minimum levels of c_1 and c_4.
3. As some decision-makers may be assigned to more than one bloc, all those decision-makers that entered a given bloc are added to blocs formed in

later steps, following the same procedure described in point 1.

The procedure does not always yield a satisfactory solution. A special option permits the iterative procedure to begin with a given bloc of two or more members, instead of starting with the minimum distance in the matrix.

Each time a decision-maker is added to a bloc the values of c_1 and c_4 of that bloc are given. After final identification of a bloc by the above procedure that bloc is analyzed with the same measures and coefficients used for intra-group analysis (see sections 4.3.1 and 4.3.2).

Although the procedure is primarily based on the average distance c_4, a minimum level of c_1 is also used, because in a larger group the distance between two decision-makers may be very high without much influencing the level of the coefficient c_4.

Mokken (1970, 190-6) proposed an analogous procedure for multiple scaling, based on the coefficient of scalability H. In Mokken's procedure, however, items from former scales are not added to scales formed in later steps (point 3), but the programs contain the possibility to do so.

4.5 *Clique analysis*

Our second procedure to define blocs or clusters of decision-makers in terms of close policy positions - clique analysis - is based on the coefficient of cohesion c_1 and the range of distances. In general it is a method of grouping elements of the basis of a dissimilarity of similarity matrix between these elements. Our procedure is based on a more generally formulated grouping method developed by Peay (1970), who makes use of concepts of graph theory to define a general method for grouping elements on the basis of a distance matrix.

A clique is defined as a maximum subset of decision-makers with the reduced distance between each pair of delegations less than the criterion distance d^r_0. In other words, a clique is a maximum subset of decision-makers with a coefficient of cohesion c_1 equal to or higher than $c_1^{(o)} = 1 + d^r_0$. Clique analysis may be distinguished from most clustering techniques in the sense that cliques do not need to be disjoint. The demand of disjointedness implies for most of the cluster procedures (also for bloc analysis) that the results of the cluster analysis are not uniquely defined by $d^r{}_0$ and the distance matrix, but depend on the starting point of the procedure (mostly the pair of decision-makers with minimal d^r).[15] The requirement that a clique must be a maximum subset of delegations implies that no clique should be a subset of another clique.

In section 4.3.2 we discussed the ordering of the group members according to their coefficient of location in the group from central to marginal. We stated that this ordering does not inform us about blocs or cliques of decision-makers in the group. Bloc and clique analysis can therefore be used well for a more detailed analysis within a priori defined groups.[16] Because of the special position of the pivotal member(s) in the group, as an intermediate position between the other group members, those cliques containing one or more pivotal member(s) are of particular interest. We call them pivotal cliques. Sometimes it may be useful to indicate the policy position of the group not only by its pivotal member(s) but also by those members that have only a small distance to the pivotal member(s), i.e. those that are contained in one or more pivotal cliques. This union of all pivotal cliques can be denoted by the concept pivotal superclique.

Which criterion $d^r{}_0$, as the maximum allowed distances in the cliques, should be used? Our choice will be based in part on the results of our analysis. An absolute criterion cannot be given. Also for bloc analysis it is impossible to give absolute criteria for the minimal levels of the coefficents c_1 and c_4.

4.6 *The treatment of absence in roll calls: abstention or missing data?*

Until now, we have been assuming a situation in which all decision-makers participated in all roll calls by choosing one of the three alternatives: in favor, against or abstention. However, decision-makers do not participate in all roll calls. What should be done in the case of nonparticipation? We have treated absence as abstention in our illustrations. In another approach, absence is considered as missing data: distances or agreements are only based on those roll calls in which both decision-makers participated (Lijphart, 1963; Baehr, 1964; Willetts, 1972, 1978). We shall now consider both possibilities.

There are a number of arguments in favor of considering absence as abstention. This equivalence of absence and abstention as one and the same outcome of the vote is based on two assumptions. First, we may argue (assumption 1) that abstence of a decision-maker A on an issue i is between that of B, voting in favor ($b_i = y$), and C, voting against ($c_i = n$). A did not choose the positive or the negative voting alternative. Only in those cases in which A was absent for reasons of principle (e.g. South Africa on apartheid questions), can absence be considered an extreme negative voting alternative. Further (assumption 2), A also would have been absent if the proposal had been put to the vote in its negative formulation. This second assumption is important in relation to the seventh axiom in section 4.1.1, the invariance for taking converses. That axiom states that the value of the distance function does not change if we substitute for vector $\vec{a}$ and $\vec{b}$ their converse factors $-\vec{a}$ and $-\vec{b}$. If we accept that absence is between voting in favor and voting against and that the converse of absence is absence, it can be deduced from our axioms that both absence and abstention should be mapped into the same numerical score, namely 0. In other words, absence should be considered as similar to abstention.

The above arguments notwithstanding, other arguments can be used to support the treatment of absence as missing data. If a decision-maker A is absent, and if the reasons for its absenteeism are unknown (a decision-maker may be ill, there may be important meetings elsewhere), we don't know how he would have voted had he been present. Particularly in large decision-making bodies like the General Assembly, with many committee meetings taking place at the same time, smaller delegations are unable to attend all meetings and therefore are unable to participate in all votes. In these instances absenteeism tells us more about the priorities of these smaller delegations than about their policy position on the issues. If we are interested in the location or policy position of such a decision-maker in relation to other decision-makers (e.g. of their caucusing groups), it would be a mistake to treat absence as abstention. Rather, we are confronted with a number of issues in which a decision-maker did not take a known position: absence is missing data.

Formulated in this way, we are confronted with a missing data problem, in which the distance between two decision-makers A and B has to be estimated on the basis of their known distance over n_{ab} roll calls, in which both decision-makers participated. Missing data are pairwise deleted, if we estimate the distance between A and B as follows:

$$d'(a,b) = \frac{n}{n_{ab}} \sum_{i=1}^{n_{ab}} | a_i - b_i |; \; 0 \leq d'(a,b) \leq 2n \quad (4.29)$$

in which n = total number of issues and n_{ab} = total number of issues in which both A and B participated. In section 4.1.2 a normalized distance $d^*(\vec{a},\vec{b})$ was defined by dividing $d(\vec{a},\vec{b})$ through its maximum value $2n$. In the same way we define $d'^*(\vec{a},\vec{b})$ as the <u>normalized distance</u> of $d'(\vec{a},\vec{b})$:

$$d'^*(a,b) = \frac{\frac{n}{n_{ab}} \sum_{i=1}^{n_{ab}} | a_i - b_i |}{2n} = \frac{\sum_{i=1}^{n_{ab}} | a_i - b_i |}{2n_{ab}} \quad (4.30)$$

$$0 \leq d'^*(a,b) \leq 1$$

A well known index of agreement for roll call analysis is Lijphart's index of agreement based on Rice (1928) and Beyle (1931). It is defined as (Lijphart, 1963, 910):

$$IA = \frac{f + \frac{1}{2} q}{t} \times 100 \quad (4.31)$$

in which: f denotes the number of similar votes of the two decision-makers (both for, both against, or both abstaining); q denotes the number of votes in which one decision-maker is voting for or against the proposal and the other is abstaining; t denotes the total number of votes in which both decision-makers participate. It can be easily shown that $d'^*(\vec{a},\vec{b})$ is a simple linear transformation of Lijphart's index of agreement:

$$IA = 100\,(1 - d'^*(a,b)) \quad (4.32)$$

Therefore, Lijphart's index of agreement is related to our distance measure and the axiomatic structure generating it, as Kendall's τ was in case of preference orders.

The proposed treatment of absence as missing data in (4.30) has, however, serious consequences for our axiomatic structure. The measure $d'(\vec{a},\vec{b})$ has no longer the desirable properties we listed in section 4.1.1 as axioms. $d'(\vec{a},\vec{b})$ possibly may violate the property of triangle inequality, which is one of the properties <u>all</u> distance functions should have. $d'(\vec{a},\vec{b})$ can therefore not be considered as a distance function. Additivity of the distances is no longer equivalent to betweenness. In other words, the metric properties of our distance function do not hold anymore, if we decide to use $d'(\vec{a},\vec{b})$ instead of $d(\vec{a},\vec{b})$ with absence equated as abstention. Neither do the statistical properties of $d^r(\vec{a},\vec{b})$ hold if $d'(\vec{a},\vec{b})$ is used. In the case of $d'(\vec{a},\vec{b})$ we can also define the corresponding reduced distance $d'^r(\vec{a},\vec{b})$ between two vote vectors $\vec{a}$ and $\vec{b}$ on a set of n roll calls:

$$d'^r(a,b) = \frac{d'(a,b)}{\mathcal{E}\{\underline{d}'(a,b)\}} \qquad (4.33)$$

However, the distribution theory of $\underline{d}'(\vec{a},\vec{b})$ is quite different from that of $\underline{d}(\vec{a},\vec{b})$. Only the expected distance $\mathcal{E}(\underline{d}'(\vec{a},\vec{b}))$ is the same as the one given in (4.10) for $\underline{d}(\vec{a},\vec{b})$. The distribution theory of $\underline{d}(\vec{a},\vec{b})$ can therefore not be used for $\underline{d}'(\vec{a},\vec{b})$ or for the coefficients in section 4.3 if they are defined in terms of $d'(\vec{a},\vec{b})$. By consequence, although the coefficients of section 4.3 can meaningfully be applied on the basis of $d'(\vec{a},\vec{b})$, they loose their statistical character.

We have presented two formal arguments for absence as abstention based on the reasonable assumptions that absence is between voting in favor and voting against and that the converse of absence is absence. The strong metric and statistical properties of our distances in this case is another argument. We may further add theoretical arguments against the treatment of absence as missing data: absence is and indicates nonparticipation in the decision-making, especially if we consider roll calls in the committees of a legislative body like the UN General Assembly, where decision-making in the plenary is prepared. Groups of members states whose delegations are well equipped to participate in all parallel activities at the same time will be able to exercise more influence in the decision-making process than groups of member states with smaller delegations. From this point of view, absence weakens the cohesion of a group and the treatment of absence data overestimates the level of cohesion as an instrument of influence.

4.7 Relation to other methods of analysis of groups and coalitions

Roll call analysis can be and has been used for a variety of research objectives: dimensions and clusters of issues can be determined; the position of individual decision-makers can be characterized on an issue or an issue dimension; *a priori* defined groups of decision-makers - such as caususing groups of, in national parliaments, parties and factions - can be analyzed, in particular in terms of voting cohesion; and dimensions and clusters of decision-makers, rather than issues, can be determined. A large number of indices and methods have been proposed for these different objectives and have been well reviewed in Anderson et.al. (1966) and MacRae (1970).

To determine cohesion and policy locations of groups of decision-makers the determination of policy positions of decision-makers relative to each other and relative to different groups of decision-makers is required. For such a pairwise comparison of the votes of decision-makers a large number of indices and measures of associations have been proposed.

One of the most well know *indices* is Lijphart's index of agreement. We have seen that it is formally related to the distance function we proposed.

A systematic survey of *measures of association* for roll call analysis was given by MacRae (1970, 41-51) and Weisberg (1968, 108-29; 1974). They only considered fourfold tables - association measures between two dichotomous variables (yes-no votes) - because the alternative "abstention" does not exist in the American Congress. The most important distinction between different types of correlation coefficients is that between measures of "one-way association" and "two-way association".

A coefficient of "*one-way association*" attains its maximum value when there is a single zero cell in the fourfold table. Yule's Q is an example of a one-way association measure often used for the analysis of roll calls. If the cells in a fourfold table are designed as

a	b
c	d

Yule's Q is given by the expression:

$$Q = \frac{ad - bc}{ad + bc} \tag{4.34}$$

Yule's Q is a special case of the Goodman-Kruskal coefficient γ (gamma), which is defined for contingency tables of any size.

One-way coefficients of association are not suitable for the analysis of cohesion of groups or coalitions of decision-makers. These coefficients measure the degree of unidimensionality between a pair of items (decision-makers or issues). They indicate that different decision-makers vote in a cumulative way. Decision-makers on extreme opposite sides of an issue dimension might well have a maximal value on a one-way coefficient of association; however, they can hardly be considered a cohesive group in such a case!

Measures of "*two-way association*" attain their maximum value if the two items (decision-makers or issues) are completely identical (two zero cells in the fourfold table above). The Pearson product-moment correlation r is the classic measure of "two-way association". These measures indicate the extent of covariation (Weisberg, 1968, 110).[17)] They are therefore more adequate for the analysis of cohesion of groups or coalitions of decision-makers. The choice of the index or coefficient of two-way association, however, should be based on a serious consideration of the *metric* properties disclosed by the data. A number of researchers neglected the metric assumptions of some measures of association, *in casu* the product-moment correlation, and factor analysis, based on these measures. Because these metric assumptions are too strong for the three- (or two-)valued types of data that we have in roll call situations, factor analysis can seriously bias the outcomes by overestimating the number of underlying dimensions (MacRae, 1970, 151-9; Weisberg, 1968, 135-9; Mueller, 1967; see also Rai, 1974). For the proper determination of the number of dimensions Weisberg and MacRae advocated factor analysis of a matrix of coefficients of "one-way association", mostly Yule's Q (MacRae, 1970, 159; Weisberg, 1968, 185-9). In fact, the factor analysis solution then approximates the different cumulative scales of issues or delegations and can therefore not be related to cohesive groups of decision-makers as we have shown above. Moreover, there is no mathematical justification for the application of factor analysis, with its metric assumptions, on such a matrix. For the detection of cumulative scales a far more elegant and analytically powerful method of multiple scaling can be used, developed by Mokken (1970). He developed a *stochastic* model of cumulative scaling and proposed a method of multiple scaling very similar to the procedure of bloc analysis we described in section 4.4. No metric assumptions are made in the model or in the procedure of multiple scaling (see also Stokman and Van Schuur, 1980).

The distance measure we introduced in section 4.1 can be considered as the two-way measure of (dis)association with the appropriate metric properties. Whereas the product-moment correlation is based on the Euclidian space, our distance measure has the properties of a *city bloc metric*, if the seven desirable properties or axioms on which it is based, seem reasonable. As sections 4.3 to 4.6 here shown, a number of analytic powerful coefficients and procedures, directly related to relative policy locations and cohesion, could be defined in terms of this distance measure.

Another index or two-way (dis)association for roll call analysis is proposed by Wolters (1976). He argues that opposing a proposal should be considered as fundamentally different from voting yes. This is particularly the case if certain no-voters oppose a proposal because it is not radical enough, whereas others oppose it because it is too radical. In these situations voting against might indicate a small distance between decision-makers as well as a very large distance. For that reason he proposes Jaccard's index, in which voting against in no way contributes to distance or similarity. Defining the four cells in the four-fold table by a, c and d, as we did above, Jaccard's index is defined as:

$$\frac{b + c}{a + b + c} \qquad (4.35)$$

This index, introduced as a distance measure, has no metric properties. In our distance measure $d(\vec{a},\vec{b})$, as defined in (4.1) both voting together yes and voting together against are neutral. As axiom 5 states the distance between two vote vectors depends *only on those roll calls in which A and B voted differently*: not voting together yes nor voting together against contain information about distances, because we do not know the reasons why decision-makers vote together. We therefore prefer our distance metric with its metric properties above that of Jaccard. Indeed, after normalization or reduction of the distances, as proposed in sections 4.1.2 and 4.1.3, voting together in favor and against loose their neutral character and contribute to a smaller distance <u>if distances over different sets of roll calls are compared with one another</u>. If this is cumbersome for some reason, <u>e.g.</u> because a large number of roll calls is supposed to exist of the nature Wolters describes, the absolute distances of (4.1) should be used for comparative analysis across sets of roll calls instead of the normalized or reduced distances. Indeed, the large number of possible models of roll-call behavior in different decision-making situations, Wolters introduces, stress the importance to attach meaning only to roll calls in which A and B vote differently.

Relations between political parties and factions in decision-making bodies might well be analyzed by application of different multidimensional scaling models. It results in a spatial representation of parties and/or decision-makers. Although such analyses might help to get an overall picture of the relations in a decision-making body, they lack coefficients to compare cohesion and policy locations. As such, these models might well be used in addition to the methods we proposed above, using our distance metric as the most adequate measure of (dis)similarity. As an example, we might mention one of the analyses on the preference rank orders of the 12 Dutch political parties by Daalder and Rusk (1972). They computed Kendall's τ between all political parties across the members of parliament. The resulting matrix of Kendall's τ's between each pair of parties was used as input for a MINISSA analysis. Our approach suggests at least two other possible multidimensional scaling analyses: one on the basis of the matrix of distances (or Kendall τ's) between *each pair of members of parliament* and one on the basis of table 4.15, containing the *inter-pivot distances*.

A final remark concerns the types of data we have considered here: roll calls and preference rank orders. These types of data are adequate to analyze (relative) *policy positions* between decision-makers. There are other aspects of group cohesion that are not directly related to policy positions, but to other functions of groups in deicision-making bodies. Particularly, one might think of the important aspects of *initiative* and *leadership* in decision-making bodies or groups of decision-makers that can not be analyzed on the basis of roll calls or preference orders. Stokman (1977) has shown that pivotal policy positions and leadership in groups of delegations within the United Nations coincide only if we consider issues that are related to the group goals of the groups concerned. Analysis of initiative and leadership requires therefore other types of data and other methodologies than we have considered here. Within the United Nations Stokman (1977) used sponsorship data for this purpose and elaborated a methodology for its analysis by combining graph theory and cumulative scaling.

Notes

1) Stokman (1977) analyzed shifts in leadership, policy positions and cohesion among developing nations over the fifties and sixties on the basis of roll calls and sponsorship of resolutions. The introduction of the distance measure and the coefficients is strongly based upon chapter 3 of that study. For the statistical aspects of our coefficients the reader is referred to that study. Moreover, it contains another methodology for the analysis of interaction and sponsorship data to determine leadership structures in a decision-making body.

2) In voting situations with only two alternatives (yes-no) the same distance function can be used. It satisfies the axioms in this case as well.

3) A system of ALGOL 60 computer programs in terms of which the various concepts and methods can be applied has been written in cooperation with the Technical Center of the University of Amsterdam. The principal programmers are W. van Hoboken, E. Smies, L. Storm, and F. van Veen.

4) $d^*(\vec{a},\vec{b})$ violates axiom 6, which stated that the minimum positive distance is 1. Note that in $d^*(\vec{a},\vec{b})$ the maximum distance is constant across sets, but the minimum positive distance varies across sets according to the number of votes; in $d(\vec{a},\vec{b})$ the minimum distance is constant (1), but the maximum distance varies according to the number of votes.

5) Stochastic variables are underlined.

6) Variances and co-variances of $\underline{d}(a,b)$, $\underline{d}(\vec{a},\vec{b})$ and $\underline{d}^r(\vec{a},\vec{b})$ are given in appendix C of Stokman (1977).

7) The choice of the probabilities for the voting alternatives will depend on the specific objectives of an analysis.

8) In Mokken and Stokman (1970, 4-4-45) two other coefficients of cohesion - $c_2(G)$ and $c_3(G)$ - were defined on the basis of the distance matrix $D(G)$ instead of $D^r(G)$. As they are not of interest here, we have deleted them

9) In fact, for a cohesive group we expect not only that $c_4(G) > 0$, but that the coefficient has too high a value to be reached in case of random voting or preference ranking. Let $c_4^{(0)}$ be the value of the coefficient under the null hypothesis of random voting or preference ranking, such that the probability of a higher value of $c_4(G)$ is less than say 5%:

$$P\{\underline{c}_4(G) \geq c_4^{(0)} \mid H^0\} \leq .05$$

For a cohesive group we should then have:

$$0 < c_4^{(0)} \leq c_4(G) \leq 1$$

To determine the value of $c_4^{(0)}$ we should know the distribution of $\underline{c}_4(G)$. For roll calls it can be proven that $c_4(G)$ is approximately normally distributed (Mokken and Stokman, 1970, 5-9-5-20). For both $n \to \sim$ and $k \to \sim$ $\underline{c}_4(G)$ has an asymptotic normal distribution. The expected value in case of random voting is $\mathcal{E}(\underline{c}_4(G)) = 0$. The variance of $\underline{c}_4(G)$ is given in appendix C of Stokman (1977). For large numbers of roll calls the variance of $\underline{c}_4(G)$ is so small, however, that already a very small deviation of 0 is significant. For practical purpose one can say that a voting group is cohesive if

$$0 < c_4(G) \leq 1$$

It was not possible to derive the variance of $c_4(G)$ for preference ranking.

10) For incoherent groups a similar line of arguments holds, as was given in note 9 for cohesive groups.

11) As in the case of $c_4(G)$ (see footnote 9) we should have for a 'loyal' group member:

$$0 < c_4^{(0)}(g_i) \leq c_4(g_i) \leq 1$$

in which $c_4^{(0)}(g_i)$ denotes the value of the coefficient under the null hypothesis of random voting, such that the probability of a higher value of $c_4(g_i)$ is less than say 5%. For roll calls $\underline{c}_4(g_i)$ has an asymptotic normal distribution for both $n \to \sim$ and $k \to \sim$. The variance of $\underline{c}_4(g_i)$ is given in Appendix C of Stokman (1977). For large numbers of roll calls the variance of $\underline{c}_4(g_i)$ is so small that already a very small deviation of 0 is significant. For practical purposes one can just say that a group member is 'loyal' if:

$$0 < c\ (g_{i\cdot}) \leq 1$$

It was not possible to derive the variance of $\underline{c}\ (g_{i\cdot})$ for preference ranking.

12) For roll calls $\underline{c}_4\ (G_1, G_2)$ has an asymptotic normal distribution for $n \to \infty$, $k \to \infty$ and $t \to \infty$. Its variance is given in Appendix C of Stokman, 1977.

13) For roll calls $\underline{c}_4\ (g_{i\cdot}, G_2)$ has an asymptotic distribution for $n \to \infty$, and $t \to \infty$. Its variance is given in Appendix C of Stokman, 1977.

14) We thank Professor Daalder (University of Leyden) for putting his data at our disposal for the purposes of these analyses.

15) In his report, Peay introduced so-called r-reachability cliques, defined as subsets of elements that are not necessarily disjoint. The elements of these subsets are connected with each other by a chain with a length or r (or less) arcs, the values of these arcs being at most equal to a criterion distance d_0 (in the case of dissimilarities). According to the general theory of Peay, our definition agrees with the so-called 1-reachability clique. The program contains the possibility of analyzing r-reachability cliques.

16) Our procedures of bloc and clique analysis are based on the same measures of voting cohesion as those used in group analysis. These different procedure give therefore directly comparable results. This is not the case with a study of Harbert (1976). He compared the voting cohesion of the mini-states in the United Nations for different issue areas over the period 1971-1972. For this purpose he modified the Rice index in such a way that abstention was taken into consideration. However, to detect different clusters among the mini-states, he used a completely different measure: Lijphart's index of agreement. By consequence, a comparison between the results of the two procedures of analysis is very difficult. Moreover, Harbert was unable to evaluate the individual contribution of the group members to the voting cohesion of the group or cluster, which we are able to do on the basis of the coefficients of location.

[17] In his article on models of statistical relationship, Weisberg classified the different measures of association on the basis of two characteristics: (1) the condition of maximum relationship, and (2) the condition of null relationship. The definitions of maximum or perfect relationship and null relationship are then combined to yield a family of different relationship models (Weisberg, 1974, 1639-43).

For the References to Chapter 4, see page 285.